(19)日本国特許庁（ＪＰ）　(12)公開特許公報（Ａ）　(11)特許出願公開番号

特開平8－320933

(43)公開日　平成８年(1996)12月３日

(51)Int.Cl.[6]	識別記号	庁内整理番号	Ｆ Ｉ		技術表示箇所
Ｇ０６Ｔ　7/00			Ｇ０６Ｆ　15/62	４１５	
Ｇ０１Ｂ　11/24			Ｇ０１Ｂ　11/24	Ｋ	

審査請求　有　請求項の数９　ＯＬ　外国語出願　（全 23 頁）

(21)出願番号　特願平8－72356

(22)出願日　平成８年(1996)３月27日

(31)優先権主張番号　０８／４１４，３９７

(32)優先日　1995年３月31日

(33)優先権主張国　米国（ＵＳ）

(71)出願人　000004237
日本電気株式会社
東京都港区芝五丁目７番１号

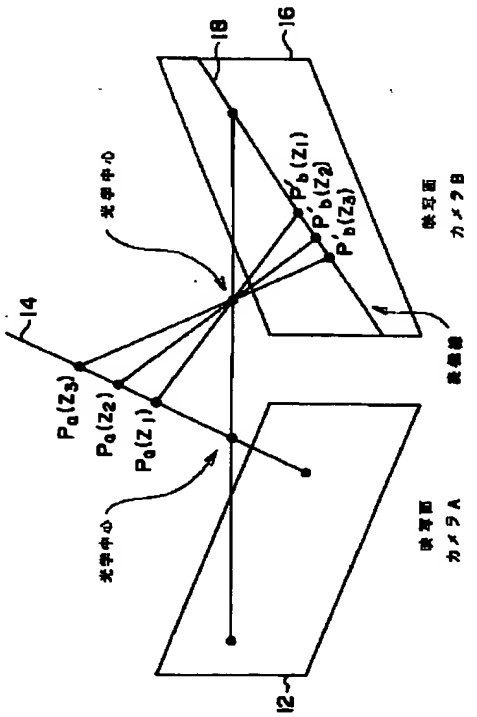
(72)発明者　インゲマー　コックス
アメリカ合衆国，ニュージャージー　08648，ローレンスヴィル，レ　パーク　ドライヴ　21

(72)発明者　セバスチャン　ロイ
アメリカ合衆国，ニュージャージー　08540，プリンストン，ムーア　ストリート　238

(74)代理人　弁理士　後藤　洋介　（外２名）

(54)【発明の名称】　三次元画像の評価方法

(57)【要約】
画像対がそのシーンの三次元画像表現を提供するために用いられているとき、１つのシーンの二次元ビューの対を記録するのに用いるカメラのエゴモーションを補償するための技術。当該技術は、強度ヒストグラムの比較、ヒストグラムの相違する四角形の総計が最小に帰着する合計を特定するためのエゴモーションの想定された合計のための画像対における対応する表極線の画素のレベルを含む。



1

【特許請求の範囲】

【請求項１】　光景に関する異なる画像対により表される視点に含まれる総回転数を知ることに依り、前記光景に関する複数の二次元画像から前記光景に関する三次元画像を得るためのプロセスにおける前記総回転数を近似する方法であって、

（ａ）前記画像対の２つの視点間におけるある特定の回転数を仮想することによって前記光景の画像対における対応する複数の表極線対を決定する工程と、

（ｂ）前記表極線のそれぞれに沿って画素密度のヒストグラムを用意する工程と、

（ｃ）前記２つの画像の対応する表極線対のそれぞれの前記ヒストグラムにおける画素密度レベルの四角形の相違の合計を決定する工程と、

（ｄ）前記四角形の相違の合計の総計を決定する工程と、

（ｅ）前記仮想回転の相異なる総計を得るために（ａ）、（ｂ）、（ｃ）及び（ｄ）の工程を繰り返す工程と、

（ｆ）前記（ｄ）工程で求めた最小計と関連した仮想回転の総計を使用する工程とを有することを特徴とする三次元画像の評価方法。

【請求項２】　前記（ａ）工程における複数の表極線対が少なくとも５０であることを特徴とする請求項１記載の三次元画像の評価方法。

【請求項３】　前記（ａ）工程は、前記仮想回転の総計の選択において、グラジエント降下の探索を用いることを特徴とする請求項１記載の三次元画像の評価方法。

【請求項４】　ヒストグラムの正規化は、第一に、画像明度におけるバリエーションを補償するために用いられることを特徴とする請求項１記載の三次元画像の評価方法。

【請求項５】　光景に関する異なる画像対により表される視点に含まれる総翻訳数を知ることに依り、前記光景に関する複数の二次元画像から前記光景に関する三次元画像を得るためのプロセスにおける前記総翻訳数を近似する方法であって、

（ａ）前記画像対の２つの視点間におけるある特定の翻訳数を仮想することによって前記光景の画像対における対応する複数の表極線対を決定する工程と、

（ｂ）前記表極線のそれぞれに沿って画素密度のヒストグラムを用意する工程と、

（ｃ）前記２つの画像の対応する表極線対のそれぞれの前記ヒストグラムにおける画素密度レベルの四角形の相違の合計を決定する工程と、

（ｄ）前記四角形の相違の合計の総計を決定する工程と、

（ｅ）前記仮想翻訳の相異なる総計を得るために（ａ）、（ｂ）、（ｃ）及び（ｄ）の工程を繰り返す工程と、

2

（ｆ）前記（ｄ）工程で求めた最小計と関連した仮想翻訳の総計を使用する工程とを有することを特徴とする三次元画像の評価方法。

【請求項６】　前記（ａ）工程における複数の表極線対が少なくとも５０であることを特徴とする請求項５記載の三次元画像の評価方法。

【請求項７】　前記（ａ）工程は、前記仮想翻訳の総計の選択において、グラジエント降下の探索を用いることを特徴とする請求項５記載の三次元画像の評価方法。

【請求項８】　ヒストグラムの正規化は、第一に、画像明度におけるバリエーションを補償するために用いられることを特徴とする請求項５記載の三次元画像の評価方法。

【請求項９】　一画像のふたつのフレームにおけるカメラの視点のエゴモーションを求めるプロセスであって、前記エゴモーションの回転成分を求めるための請求項１記載の方法及び前記エゴモーションの翻訳成分を求めるための請求項５記載の三次元画像の評価方法。

【発明の詳細な説明】

【０００１】発明の属する技術分野

本発明は、コンピュータビジョン、より詳しくは、光景に関する複数の二次元画像から前記光景に関する三次元画像を抽出するためのコンピュータの利用乃至は対象物の相異なる視点の方向における変化を知ることによる他の利用方法に関する。

【０００２】従来の技術

光景乃至は対象物に関する二次元画像から光景乃至は対象物に関する三次元画像を再構築するための探索を行うコンピュータビジョンシステムにおいて、重要なパラメータは、当該光景の相異なる視野の視点における変化である。当該光景の二つの画像が、例えばノイズから生じるような、エゴモーションと呼ばれる当該光景を記録するカメラの未知の回転及び翻訳を含む二つの視野を表しているとき、三次元画像を忠実に再構築するためには、かなりのコンピューティングが含まれる。三次元画像の忠実な再構築は、例えば、ナビゲーションにおける移動量の評価、ビデオのモザイクがけ、対象物の二つの二次元画像から三次元画像を抽出すること、光景の相違する部分の多くの視野を積分して光景全体の一つの視野にすること、等多くの応用に有効であることが、ＡＲＡＰイメージアンダースタンディングワークショップ１９９４のＰｒｏｃにおける“マルチビューからの形状の回復：パララックスベースのアプローチ”と題されたＲ．Ｋｕｍａｒ等の文献に記載されている。

【０００３】一つの光景の二つの画像フレームからエゴモーションと構造フォームを評価する問題は、コンピュータビジョンにおいて長く研究されてきた。初期の頃から構造と動きのアルゴリズムを二つの明瞭な分類に分ける試みがあった。第一は、フィーチャベースのものであり、これは、前記の二つの画像フレーム間には既知の数

のフィーチャ対応があると仮定するものである。理論的には、構造と動きの問題を解決するためにフィーチャ対応はほとんど必要無いが、このアプローチは大変ノイズに敏感であり、安定した解を得るためには実際多くの対応が必要である。更に、フィーチャ対応を一概には認識できず、これらを発見するのが困難であるケースがしばしばである。第二のアプローチは、正確なフィーチャ対応は必要でない構造と動きの評価の直接的方法の分類を含む。このアプローチを用いた解法は、広く二つの主なサブクラスに分類され得る。問題にアプローチする一つのサブクラスは、第一に、含まれるフレームの光の流れ場の知識を開発することである。第二のサブクラスは、コンピュータビジョン、第2巻、１９８８、５１－７６頁における“動き回復の直接的方法”と題されたＢ．Ｋ．Ｐ．ＨｏｒｎｅとＥ．Ｊ．Ｗｅｌｄｏｎ，Ｊｒの文献に記載されているように、構造と動きの解を直接開発するための明度変化の束縛式を開拓することである。

【０００４】発明の概要

本発明は、その光景と関連付けた三次元回転空間内を探索することに基づいた光景の一対の二次元画像又はカメラフレーム間の回転するエゴモーションを評価するための直接的方法を含む。この方法は、各エゴモーションが相互に関連して評価され得るような画像のプロパティが存在し、その結果特定のエゴモーションが三次元画像表現に用いるのに最適なものとして特定され得る場合にのみ可能である。

【０００５】本発明の特徴は、対応すると想定され得る表極線に沿って計算される画素密度のヒストグラムのプロパティを新規に使用したことである。これらの有用なプロパティは、第１に一定の画像明度の仮定に依存し、その結果対応する表極線のヒストグラムはかみ合わない（ふさぐのを無視）こと及び略対応する表極線のヒストグラムは類似しており、この類似は当該画像における空間的相関を提供する機能となることを想定させる。画像明度におけるバリエーションを補償するために用いられ、それによって上記想定を満足させ得るヒストグラムの正規化のような有用な技術がある。

【０００６】２つの表極線の２つのヒストグラム間の相違は２つの表極線が真に対応する時に最小であり、２つの表極線間の誤整列の度合いに従って増加する、というプロパティは、三次元表極線探索として通常の態様で評価されるように２つの表極線間の回転運動を許す。

【０００７】従って、離間して配された２つの視点から得られる同一のシーンの２つのカメラフレーム間の合計は、以下のように、有効に評価され得る。第１に、純粋な回転の総計は視点の相違に含まれると仮定され、その仮定に基づいて、既知の方法によって２つのフレームのために表極線が引きだされる。各フレームのために、たくさんの対応する表極線に沿った画素密度のヒストグラムが引きだされる。２つのフレームから、選択された数の表極線それぞれのために対応する表極線のヒストグラム間の四角形の相違の合計が引きだされ、これは回転の合計を特に想定するための有用な形状として役立つ。このプロセスは回転の合計が相違するだけ繰り返され、また適当な探索、例えば、グラジエント降下又はピラミッド化が、有用な形状の最低値を提供する想定された回転を発見するために実行される。そのように想定された回転の合計は、シーンの三次元画像表現を引き出すためにフレームの以後のプロセスにおいて実際の回転の合計として扱われる。２つの視点の分離又は翻訳が重要である段階において、回転の合計を求めるために上記手続きを繰り返すことによって、分離又は翻訳の合計を近似することが望ましい。第１に翻訳を評価し、その後にエゴモーションの回転を評価するのが好ましい。

【０００８】本発明は、添付図面と共に以下の詳細な説明からより良く理解されるであろう。

【０００９】発明の詳細な説明

本発明の実施形態を詳細に説明する前に、図１の助けを借りて表極線幾何学におけるいくらかの背景を提供するのが助けとなろう。この目的のため、まず、１つのシーンのわずかに相違するビュー間の表極線関係を述べる単純な数学を概観していただきたい。斜め映写で、カメラＡ（図示せず）の映写面１２における下記の数１式に示す映写点は、異なる深さＺａを持つ三次元の点Ｐａ（Ｚａ）の線１４の映写になり得る。

【００１０】

【数１】

$$P'_a = [x'_a \; y'_a \; 1]^T$$

また、以下の数２式が導かれる。

【００１１】

【数２】

$$P_a(z_a) = \begin{bmatrix} x'_a z_a / f \\ y'_a z_a / f \\ z_a \\ 1 \end{bmatrix}$$

ここにｆは焦点距離、カメラＢ（図示せず）の映写面１６に上記諸点を映写することは、表極線１８を構成する下記数３式に示す一群の点を提供する。

【００１２】

【数３】

$$P'_b(z_a) = [x'_b \; y'_b \; 1]^T$$

また、以下の数４式が導かれる。

【００１３】

【数４】

5　　6

$$P_b^*(z_a) = J \cdot P_b(z_a) = J \cdot T_{ab} \cdot P_a(z_a) - \begin{bmatrix} f(z_a A + t_{14}) \\ f(z_a B + t_{24}) \\ z_a C + t_{34} \end{bmatrix}^*$$

ここにＴａｂは二つのカメラ間のコーディネイト変形を表す。

【００１４】また、以下の数５式が導かれる。

【００１５】

【数５】

$$\begin{bmatrix} A \\ B \\ C \end{bmatrix} - \begin{bmatrix} t_{11} & t_{12} & t_{13} \\ t_{21} & t_{22} & t_{23} \\ t_{31} & t_{32} & t_{33} \end{bmatrix} \cdot \begin{bmatrix} x_a'/f \\ y_a'/f \\ 1 \end{bmatrix}$$

ここにｔｉｊは行列Ｔａｂの要素（ｉ，ｊ）である。そして、映写行列Ｊは以下の数６式で定義される。

【００１６】

【数６】

$$J - \begin{bmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}$$

$$\vec{M}_{P_a} - \lim_{z_a \to \infty} P'_a - \begin{bmatrix} fA/C - x_a' \\ fB/C - y_a' \\ 1 \end{bmatrix}^* \quad (1)$$

また、第２の要素は表極線ベクトル、又は変位の翻訳部分であり、以下の数１０式で定義される。

※【００２０】

※【数１０】

$$\vec{E}_{P_a} - P_b'(z_{min}) - \lim_{x_a \to \infty} P_b'(z_a)$$

$$= \begin{bmatrix} f(t_{14}C - t_{24}A)/C(t_{34} + C z_{min}) \\ f(t_{24}C - t_{34}B)/C(t_{34} + C z_{min}) \end{bmatrix} \quad (2)$$

ここに、ＺｍｉｎはＰａに期待される最小深さである。これらの要素は以下の数１１式で定義される単純な関係を引き出すために用いられる。

★【００２１】

【数１１】

$$P_b' - P_a' + \vec{M}_{P_a} + e\vec{E}_{P_a} \quad 0 \le e \le 1 \quad (3)$$

ここに、ｃは表極線ベクトルに沿った不均衡である。方程式（１）（２）は回転の変位は独立した距離であり、翻訳の変位は、図２に示すように、距離に逆比例する合☆50

＊また、以下の数７式が成り立つならば、数７式で定義されるＰ　は点Ｐの映写コーディネイト表現であり、Ｐｔは以下の数８式で定義される。

【００１７】

【数７】

$$P, \ If P^* = [u\ v\ w]^T$$

【００１８】

【数８】

$$P \text{ is } [u/w\ v/w\ 1]^T$$

画像点Ｐ＇ａの変位は、二つの要素に分解され得る。第１の要素は変位の回転部分であり、以下の数９式で定義される。

【００１９】

【数９】

☆計によって表極線に沿って点をシフトさせる。

【００２２】表極線幾何学のより詳細な議論はＰＲＯＣ．ＩＥＥＥ　３ｒｄ　ワークショップ　オン　コンピ

ュータビジョンレプリゼンテーションアンドコントロール、第２巻、１９８５、１６８－１７８頁における“表極線画像解析：動きシーケンスの解析技術”と題されたＲ．Ｃ．ＢｏｌｌｅｓとＨ．Ｈ．Ｂａｋｅｒの文献に記載されている。この背景をもって、本発明の理論的基礎を置くことができる。

【００２３】先に述べたヒストグラムの第１のプロパティと一致するように、一定の明度の束縛の適用、即ち、カメラの動きによっても画像点の明度が不変であるということ（１）及びかみ合いの数は小さいということ（２）を仮定するならば、２つの表極線は本質的に等しい画素密度を含み、それらの位置は深さによってのみ変わるので、２つの対応する表極線の密度ヒストグラムは同一になるということは明らかである。

【００２４】図２に示すように、その回転あるいは翻訳成分のいずれかにおいてカメラの動きが小さな変化を含むケースを考える。その結果、方程式（３）における表極線は、誤りとなるが、真の表極線に近いものとなる。

【００２５】このことは、先に述べたヒストグラムの第２のプロパティの使用を用意させる。

【００２６】一定の明度の束縛の適用（１）及びかみ合いの数は小さいということ（２）を仮定するならば、一対の真に対応する表極線に空間的に近い２つの疑似表極線の強度ヒストグラムは同様の（四角形の相違の合計という意味）ヒストグラムを持つ。２つの疑似表極ヒストグラム間の相違は、表極線が真の表極形状に対応し、回転誤差のサイズと共に漸次略増加する時、最小である。このプロパティが自然画像に一般的に適用されることは、以下のように、演繹される。画像強度は空間的に高い相関を有することは良く知られている。図３に述べるように、カメラの変位Ｔａｂにおける小さな誤差は、画像Ａにおける点Ｐ＾ａを下記数１２式で示される真の表極線に空間的に近い点に映写させる。

【００２７】

【数１２】

$$\vec{E}_{P_a}$$

上記誤差が小さくなればなるほど、この点は下記数１３式で示される点に近くなる。

【００２８】

【数１３】

$$\vec{E}_{P_a}$$

局所の画像を結合させることは、誤差対応の強度値が真の表極線のどこかに存在する真の強度値に近くなることを保証する。

【００２９】第２のプロパティを保持しない人工画像を構成することは容易であるが、これらは決して自然のものではない。例えば、回転不変の円の画像は、回転のｚ成分を評価させない。しかしながら、一般にこのプロパティは画像の多くに保持されていると信じられている。

【００３０】回転誤差と翻訳誤差の両結果（それぞれ図３Ａ、図３Ｂ）を比較することによって、翻訳誤差は、通常、真の表極線からの変位は回転誤差のそれよりも少ない。翻訳誤差からの変位の大きさは、当該シーンにおける対象物の最小深さによって、“逆スケール”であるが、回転誤差からの変位の大きさはそうではない（方程式（１）と（２）を見よ）。

【００３１】これは、対象物が近すぎないならば、回転誤差は常に翻訳誤差よりも大きなインパクトを持つことを意味する。全対象物が（無限の）背景にあるような限界事例では、翻訳誤差は全く変位を生じない。

【００３２】上記相関から重要な結論が引き出せる。翻訳誤差が通常生じる真の表極線からの変位の合計は“無視”し得るものである。このように、通常の事例では、すべての点の変位は回転誤差により生じる。後で、通常の事例について、適当なアプローチを試みる。

【００３３】この理論的背景を下に、本発明の方法について、以下に記載する。

【００３４】図４は、離間して配された２つのカメラ又はその２つのフレームを記録するために動かされた１つのカメラのいずれかにより撮られたひとつのシーンの２つのフレーム間の未知の回転の合計を求めるための方法を示すフローチャートである。この目的のため、より典型的な事例であるが、一定の画像明度を想定する。フローチャートで述べられるように、第１のステップ４１は、回転の近似値を想定し、これを元に、２つのフレームの対応する表極線を引き出す。典型的にはフレームの表極線の少なくとも４分の１、好ましくは当該フレームで用いられた略全表極線というように、沢山のそのような表極線が引き出される。ノイズへの敏感さを減じるという理由から、引き出される表極線の数が増えれば、正確さも増す。そして、第２のステップ４２として、２つのフレームの選択された対応する表極線対に沿って、画素の明暗度のヒストグラムが用意される。続いて、第３のステップ４３として、対応する表極線対それぞれのために、そのような表極線対のヒストグラムから四角形の相違の合計が分離抽出される。そして、ステップ４４として、すべての表極線対の四角形の相違の合計を総計したものが、回転の合計の有用な形状に用いるために求められる。このプロセスが、想定された相違する回転の合計ごとに繰り返される。第２の有用な形状が第１のものよりも小さいならば、想定されたより大きい回転の合計まで、繰り返される。第２の有用な形状が第１のものよりも大きいならば、元の合計よりも小さい想定された回転の合計まで、繰り返される。グラジエント降下探索でも同様に、このプロセスが、有用な形状の最小あるいは最小近くで歩留まる回転が見出されるまで繰り返される。そのような最小で歩留まる回転の合計は、本質的に回転の真の合計である。一旦回転の合計が既知となれ

ば、これは１つのシーンの２つのフレームに関する既知の態様に当該シーンの極めて正確な３次元画像表現を構成するために用いることができる。

【００３５】代替的には、まず適当な値を発見するために粗い探索を始め、その後、前の探索で境界を定めた狭い領域をより細かく細かく探索していくピラミッド化探索が利用できる。

【００３６】当該画像が一定の画像明度の想定を満足することを保証するために、２つの画像は、まずヒストグラム正規化のプロセスによって正規化される。このことは、コンピュータビジョン＆パターンレコグニション（１９９４）、７３３－７３９頁における“最大近似されたNカメラステレオアルゴリズム”と題されたＩ．Ｊ．Ｃｏｘの文献に、あるいは“ディジタル画像処理”と題されたＧｏｎｚａｌｅｚ　ａｎｄ　Ｗｉｎｔｚの文献に掲載されたも下グラム仕様に記載されている。

【００３７】図４は本発明により実施されるプロセスのフローチャートとして記載されているが、プロセス実行のために設計された装置のハードウエア構成のブロックダイアグラムとしても役立つ。特に、各ブロックは、そのための動作ステップを実行するために設計された特別目的のコンピュータとなり得る。

【００３８】上記手続きで前に述べたように、２つのビューにおけるカメラのいかなる翻訳動作も、回転動作を求める上で無視し得る結果を有するとして考慮しなくてよい。例えば、動作は完全に１つのタイプ、例えば、回転及びそのような回転動作の近似を引き出すために論じられる態様への進行であると想定して始めてもよい。これは、翻訳動作の近似を得るために、そのような動作の固定値として発見された回転の近似を用いて、同一の一般的アプローチを利用することで続行できる。一旦回転動作が既知となれば、翻訳動作を評価する有用な技術がある。例えば、特に高い精度が望まれる時には、過去の発見された翻訳動作の近似を用いて回転動作の改良された近似を引き出すことで、回転動作の新らしい近似を引き出せる。この継承的近似の方法により非常に高い精度が得られる。

【００３９】対象物に関する複数の二次元画像から三次元画像を構成することは、ＭＩＴ出版、ケンブリッジ、マサチューセッツ（１９９３）発行の“三次元コンピュータビジョン”と題されたＯｌｉｖｅｒ　Ｆａｕｇｅｒａｓの本の第６章ステレオビジョン、１６５－２４０頁に記載されている。

【００４０】記載された特定の実施例は本発明の原理の暗示であることが理解されよう。対象物又はシーンの異なるフレーム間に含まれるカメラの翻訳又は回転の合計を知ることは重要である分野に本発明の原理は拡張できる。例えば、シーンの連続的なフレームを撮る乗り物又はロボットに装着されたカメラが、その位置とカメラの回転又は翻訳を知るために過去のシーンを動かすことが重要となるナビゲーションへの応用がある。

【図面の簡単な説明】

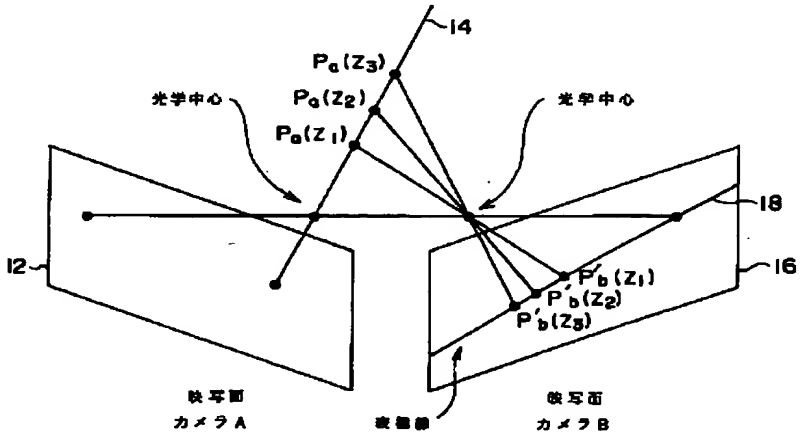
【図１】表極線幾何学の事前の議論の助けとなる。

【図２】カメラの動きによって生じる表極線における変位の回転及び翻訳成分を示す。

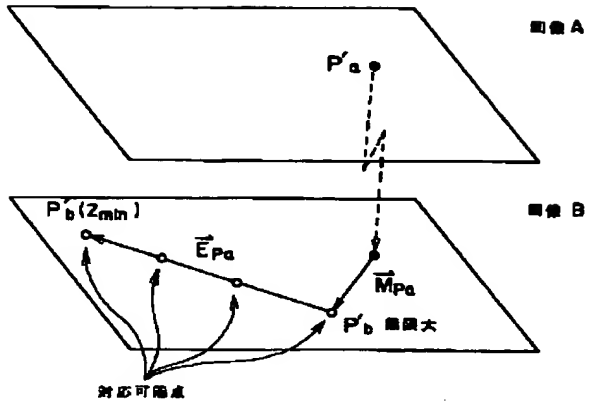
【図３】不正確な翻訳と不正確な回転のための表極線における誤差を示す。

【図４】本発明で用いられる基本的な手順のフローダイアグラムを示す。

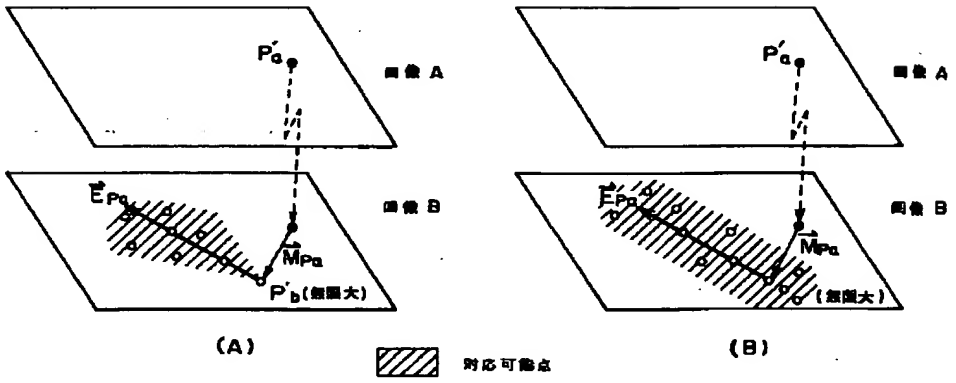
【図１】



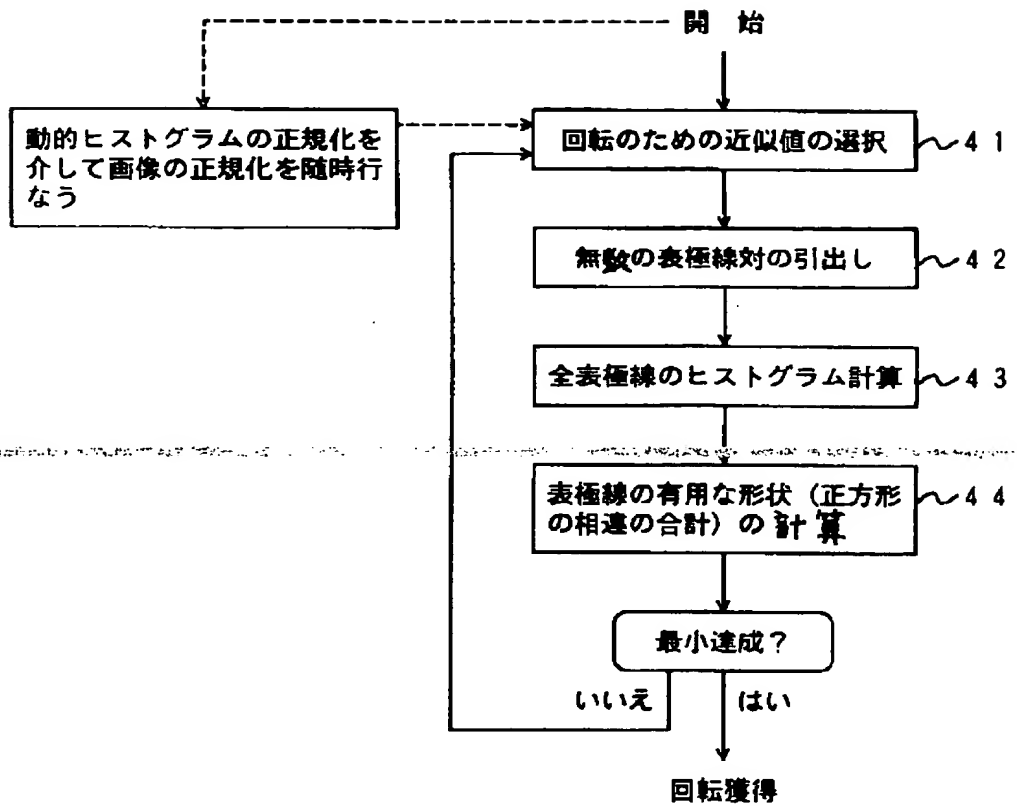
【図2】


画像A
P'a
P'b(Zmin)
画像B
EPa
MPa
P'b 最適大
対応可能点

【図3】


画像A
P'a
画像B
EPa
MPa
P'b(最適大)
(A)
(B)
対応可能点

【図4】


開　始
動的ヒストグラムの正規化を介して画像の正規化を随時行なう
回転のための近似値の選択
41
無数の表極線対の引出し
42
全表極線のヒストグラム計算
43
表極線の有用な形状（正方形の相違の合計）の計算
44
最小達成？
いいえ
はい
回転獲得

【外国語明細書】

1. Title of Invention

A method for the estimation of a three-dimensional representation

2. Claims

1. In a process for the three dimensional representation of a scene from a plurality of two-dimensional images of the scene that depends on knowing the amount of rotation involved in the viewpoints represented by a pair of different images of the scene, the method for approximating the amount of rotation involved comprising the steps of:

   (a) determining a plurality of corresponding pairs of epipolar lines in a pair of images of the scene assuming a specific amount of rotation between the two viewpoints of the pair of images;

   (b) preparing a histogram of the pixel intensities along each of the epipolar lines;

   (c) determining the sum of the squared differences of the pixel intensity levels of the histograms of each pair of corresponding epipolar lines of the two images;

   (d) determining the total of such sums;

   (e) repeating steps a, b, c and d for different amounts of assumed rotation; and

   (f) using the amount of assumed rotation that is associated with the smallest total determined in step d.

2. The method of claim 1 in which the plurality of pairs of epipolar lines in step a is at least fifty.

3. The method of claim 1 in which step a uses a gradient descent search in the choice of the amount of the assumed rotation.

4. The method of claim 1 in which histogram normalization is first used to compensate for variations in image brightness.

5. In a process for the three dimensional representation of a scene from a plurality of two-dimensional images of the scene that depends on knowing the amount of translation involved in the viewpoints represented by a pair of different images of the scene, the method for approximating the amount of translation involved comprising the steps of:

   (a) determining a plurality of corresponding pairs of epipolar lines in a pair of images of the scene assuming a specific amount of translation between the two viewpoints of the pair of images;

   (b) preparing a histogram of the pixel intensities along each of the epipolar lines;

   (c) determining the sum of the squared differences of the pixel intensity levels of the histograms of each pair of corresponding epipolar lines of the two images;

   (d) determining the total of such sums;

   (e) repeating steps a, b, c and d for different amounts of assumed translation; and

   (f) using the amount of translation assumed that is associated with the smallest total determined in step d.

6. The method of claim 5 in which the plurality of pairs of epipolar lines in step a is at least fifty.

7. The method of claim 5 in which step a uses a gradient descent search in the choice of the amount of the assumed translation.

8. The method of claim 5 in which histogram normalization is first used to compensate for variations in image brightness.

9. In a process for determining the egomotion of the viewpoint of a camera in two frames of an image, the process of claim 1 for determining the rotational component of the egomotion and the process of claim 5 for determining the translational component of the egomotion.

3. Detailed Description of Invention

Field of the Invention

This invention relates to computer vision and more particularly to use of a computer to develop a three-dimensional representation of a scene from two-dimensional representations of the scenes and other uses that depend on knowledge of changes in orientation of different views of an object.

Background of the Invention

In computer vision systems that seek to reconstruct a three-dimensional representation of a scene or object from two-dimensional images of the scene or object, important parameters are the changes in viewpoints of the different views of the scene. When two images of the scene represent two views that involve unknown rotation and translation of the camera recording the scene, to be termed ego-motion, such as might result from noise, considerable computation is involved in making a faithful three-dimensional reconstruction. A faithful three-dimensional reconstruction has utility in many applications, such as estimation of travel in navigation, three-dimensional representation of an object from two two-dimensional representations and video mosaicing, the integration of many views of different parts of a scene into a single view of the total scene, such as is described in an article by R. Kumar et al entitled, "Shape recovery from multiple views: a parallax based approach," in the Proc. of ARAP Image Understanding Workshop, 1994.

The problem of estimating the ego-motion and structural form from two image frames of a scene has long been studied in computer vision. There have been primarily two distinct classes of structure-and-motion algorithms that have been tried. The first is feature-based and assumes that there is a known number of feature-correspondence between the two frames. While few correspondences are needed in theory to solve the structure-and-motion problem, this approach is very sensitive to noise and many correspondences are in fact needed to stabilize the solution. Moreover, it is often the case that no feature-correspondences are known a priori and finding these can be laborious.

The second approach involves a class of direct methods of motion-and-structure estimating in which explicit feature-correspondences are not required.

Solutions using this approach can be broadly categorized into two main subclasses. One subclass approach to the problem is first to develop knowledge of the optical flow field of the frames involved. The second subclass approach has been to exploit the brightness-change constraint equation directly to develop solutions for motion and structure, as is described in an article by B.K.P. Horne and E.J. Weldon, Jr. entitled, "Direct Methods for Recovering Motion," in Int. J. of Computer Vision, vol. 2, 1988, pages 51–76..

Summary of the Invention

The present invention involves a direct method for estimating the rotational ego-motion between a pair of two-dimensional images or camera frames of a scene that is based on a search through the three-dimensional rotational space that is associated with the scene. This is possible if, and only if, there exists image properties such that each hypothesized ego-motion can be evaluated relative to one another so that a particular ego-motion can be identified as the most appropriate one for use in the three-dimensional representation.

A feature of the invention is the novel use of the properties of intensity histograms computed along epipolar lines that can be supposed to be corresponding. These useful properties first depend on the assumption of constant image brightness so that one can assume that the histograms of corresponding epipolar lines are invariant (ignoring occlusions) and that the histograms of almost corresponding epipolar lines are similar, this similarity being a function of the spatial correlation present in the image. There are available techniques such as histogram normalization that can be used to compensate for variations in image brightness and thereby satisfy the assumption.

The property that the difference between two histograms of two epipolar lines is a minimum when the two epipolar lines truly correspond and increases monotonically with the degree of misalignment between two epipolar lines allows the rotational motion between the two to be estimated in a straightforward manner as a three-dimensional epipolar search.

Accordingly, the amount of rotation between two camera frames of the same scene taken from two viewpoints that are spaced apart can be effectively estimated as follows. First, there is assumed that a certain amount of pure rotation was involved in the difference in viewpoints and based on such assumption there are derived epipolar lines for the two frames by known methods. For each frame, histograms of the pixel intensities along a number of corresponding epipolar lines are derived. There is then derived the sum of squared differences between the histograms of corresponding epipolar lines from the two frames for each of the chosen number of epipolar lines of the two frames and this serves as a figure of merit for the particular assumption of the amount of the rotation. This process is repeated with different assumed amounts of rotation and a suitable search, for example gradient descent or pyramidal, is carried out to find the assumed rotation that gives the lowest value of the figure of merit. The amount of rotation of such assumption is then treated as the actual amount of the rotation in the further processing of the frames to derive three-dimensional representations of the scene involved or other uses. In instances where the separation or translation of the two viewpoints may be significant, it may be desirable to approximate the amount of such separation or translation by repeating above the procedure or other suitable procedure using instead assumptions as to the separation either after or before the above procedure for determining the amount of rotation. In some instances, it may be preferable first to estimate the translation and thereafter to estimate the rotation of the ego-motion.

The invention will be better understood from the following more detailed description taken with the accompanying drawing.

Detailed Description of the Invention

Before discussing in detail the practice of the invention, it will be helpful to provide some background in epipolar geometry with the aid of FIG. 1. To this end we begin with a brief review of some simple mathematics to describe the epipolar relationship between two slightly different views of a scene. With perspective projection, a projected point $P'_a = [x'_a\ y'_a\ 1]^T$ in projection plane 12 of camera A (not shown) can be the projection of a line 14 of three-dimensional points $P_a(z_a)$ of different depth z_a. We then have

$$P_a(z_a) = \begin{bmatrix} x'_a z_a / f \\ y'_a z_a / f \\ z_a \\ 1 \end{bmatrix}$$

where f is the focal length. Projecting those points to the projection plane 16 of camera B (not shown) gives a set of collinear points $P'_b(z_a) = [x'_b\ y'_b\ 1]^T$ that will form the epipolar line 18.

We also have

$$P^*_b(z_a) = J \cdot P_b(z_a) = J \cdot T_{ab} \cdot P_a(z_a) = \begin{bmatrix} f(z_a A + t_{14}) \\ f(z_a B + t_{24}) \\ z_a C + t_{34} \end{bmatrix}^*$$

where T_{ab} represents the coordinate transformation between the two cameras and

$$\begin{bmatrix} A \\ B \\ C \end{bmatrix} = \begin{bmatrix} t_{11} & t_{12} & t_{13} \\ t_{21} & t_{22} & t_{23} \\ t_{31} & t_{32} & t_{33} \end{bmatrix} \cdot \begin{bmatrix} x'_a / f \\ y'_a / f \\ 1 \end{bmatrix}$$

and t_{ij} is element (i, j) of matrix T_{ab}. The projection matrix J is defined as

$$J = \begin{bmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}$$

and P^* is the projective coordinate representation of a point P. If $P^* = [u\ v\ w]^{*T}$, then the homogeneous euclidean coordinate P' is $[u/w\ v/w\ 1]^T$.

The displacement of the image point P'_a can be decomposed into two components. The first component is the rotational part of the displacement and is defined as

$$\vec{M}_{P_a} = \lim_{z_a \to \infty} P'_a = \begin{bmatrix} fA/C - x'_a \\ fB/C - y'_a \\ 1 \end{bmatrix} \quad (1)$$

while the second component is the epipolar vector, or translational part of the displacement, and is defined as

$$\vec{E}_{P_a} = P'_b(z_{min}) - \lim_{z_a \to \infty} P'_b(z_a) = \begin{bmatrix} f(t_{14}C - t_{34}A)/C(t_{34} + Cz_{min}) \\ f(t_{24}C - t_{34}B)/C(t_{34} + Cz_{min}) \end{bmatrix} \quad (2)$$

where z_{min} is the minimum depth expected for P_a. Those components can be used to derive the simple relation

$$P'_b = P'_a + \vec{M}_{P_a} + e\vec{E}_{P_a}, \quad 0 \leq e \leq 1 \quad (3)$$

where e is the disparity along the epipolar vector. Equations (1) and (2) indicate that the rotational displacement is independent of distance while the translational displacement shifts points along the epipolar line by amounts that are inversely proportional to distance, as illustrated in FIG. 2.

A more detailed discussion of epipolar geometry is provided by a paper entitled "Epipolar-Plane Image Analysis: A Technique for Analyzing Motion Sequences" by R.C. Bolles and H.H. Baker that appeared in PROC. IEEE 3rd Workshop on Computer Vision Representation and Control, pp. 168–178, (1985) and such paper is incorporated herein by reference. With this background, we can lay a theoretical basis for the invention.

Consistent with the earlier mentioned first property of histograms, if we assume (1) that the constant brightness constraint applies, i.e. the brightness of an imaged point is unchanged by the motion of the camera, and (2) that the number of occlusions is small, then it is clearly the case that the histograms of the intensities of two corresponding epipolar lines are identical since the two lines contain essentially identical pixel intensities, only their position may be changed because of depth.

Now we consider the case in which the camera motion contains a small change, either in its rotational or translational component, as represented in FIG. 2. As a consequence, the "epipolar" lines of Equation (3) above will be erroneous, but close to the true epipolar lines.

This now prepares us for use of the second property of histograms mentioned earlier. Assuming (1) that the constant brightness constraint applies, and (2) that the number of occlusions is small, then the intensity histograms of two "pseudo-epipolar" lines that are spatially close to a pair of truly corresponding epipolar lines have similar (in a sum of squared errors sense) histograms. The difference between two pseudo-epipolar histograms is a minimum when the lines correspond to the true epipolar geometry and increases approximately monotonically with the size of the rotational error.

That this property applies generally to natural images can be deduced as follows. It is well known that image intensities are spatially highly correlated. As depicted in FIG. 3, small errors in the camera displacement T_{ab} cause a point P'_a in image A to be projected to a point which is spatially close to the true epipolar line $\vec{E}_{P_a}$. The smaller the error, the closer this point is to $\vec{E}_{P_a}$. Local image coherence then insures that the intensity value of an erroneous correspondence is close to the true intensity value that lies somewhere on the true epipolar line.

While it is easy to construct artificial images for which the second property does not hold, these images are never natural. For example, an image of a rotationally invariant circle would not allow the z component of rotation to be estimated. However, in general, we believe this property to hold for a large class of images.

By comparing the effects of translational error and rotational error, (FIG. 3A and FIG. 3B, respectively), it can be shown that translational error usually creates less displacement from the true epipolar line than rotational error. This is due to the fact that the displacement magnitude from translational error is "inversely scaled" by the minimum depth of the objects in the scene, while the displacement from rotational error is not (see Equations (1) and (2)).

This implies that if the objects are not too close, the rotational error always has a much bigger impact than translational error. In the limit case where all objects are in the background (at infinity), the translation error does not create any displacement at all.

One can derive an important conclusion from this relation. The translational error generally creates a "negligible" amount of displacement from the true epipolar line. Thus one can assume in the usual case that rotational error causes all point displacement. There will be discussed later a suitable approach for the unusual case.

With this theoretical basis as a background, we can now proceed to a description of the process of the invention.

FIG. 4 is a flow chart of the process for determining the unknown amount of rotation between two frames of a scene taken either by two cameras that are spaced apart or one camera that has been moved to record the two frames. For this process, constant image brightness, which is the more typical case, is being assumed. As depicted in the flow chart, the first step 41 is to assume a likely value for the rotation and on this basis derive corresponding epipolar lines of the two frames. One would derive a number of such lines, typically at least one quarter of the lines in the frame and preferably about as many lines as were used in the frame, the accuracy generally improving the greater the number, because of the reduced sensitivity to noise this achieves. Then, as a second step 42, there are prepared histograms of the pixel intensities along the selected pairs of corresponding epipolar lines of the two frames. Then, as a next step 43, for each of the pairs of corresponding epipolar lines, in turn there is separately derived from the histograms of such pairs of lines the sum of squared differences. Then, as step 44 the total of these sums of squared differences for all of the pairs is determined for use as a figure of merit of the assumed amount of rotation. The process is then repeated to derive a figure of merit for a different assumed amount of rotation. If the second figure of merit is smaller than the first, the process is repeated with a still larger assumed amount of rotation. If the second figure of merit was larger than the first, the process is repeated with an assumed amount smaller than the original amount. In similar fashion in a gradient-descent search, the process is repeated until one finds the rotation that yields the minimum or near minimum of the figure of merit. The amount of rotation that yielded such minimum is essentially the true amount of the rotation. Once the amount of rotation is known, this can be used in known fashion in conjunction with the two frames of the scene to construct a quite accurate three dimensional representation of the scene.

Alternatively, a pyramidal search can be used in which one begins with a coarse search to find an approximate value and to follow it up with finer and finer searches centered about the narrowed region delimited by the previous search.

In order to ensure that the images satisfy the constant image brightness assumption, the two images can be first normalized by a process of histogram normalization, which is described in an article by I.J. Cox entitled "A Maximum Likelihood N-Camera Stereo Algorithm," published in the proceedings of the Int. Conf. Computer Vision & Pattern Recognition (1994), pages 733–739, or histogram specification, which is described in an article by Gonzalez and Wintz entitled "Digital Image Processing."

It can be appreciated that while FIG. 4 has been described as a flow chart of the process practiced by the invention, it can also serve as a block diagram of hardware components of apparatus designed to carry out the steps that are set forth. In particular, each of the blocks could be a special purpose computer designed to carry out the operating step prescribed for it.

As was previously mentioned in the above procedure, there has been assumed that any translational motion of the camera in the two views could be ignored as having a negligible effect on determining the rotational motion. In some instances, one may begin by assuming that the motion is entirely of one type, for example rotational, and proceed in the manner discussed to derive an approximation of such rotational motion. This could then be followed by use of the same general approach, using the rotational approximation found as the fixed value of such motion, to get an approximation of the translational motion. There are available techniques for estimating the translational motion once there is known the rotational motion. In instances when especially high accuracy is desired, there can now be derived a new approximation of the rotational motion, using the last discovered approximation of the translational motion to derive an improved approximation of the rotational motion. In this fashion by successive approximations, a very high degree of accuracy should be obtainable.

The construction of a three dimensional representation of an object from a pair of two-dimensional representations of the object is described in Chapter 6, Stereo Vision, pps. 165–240 of a book entitled "Three-Dimensional Computer Vision" by Oliver Faugeras published by the MIT Press, Cambridge Massachusetts (1993).

It should be understood that the specific embodiments described are illustrative of the general principles of the invention. In particular it should be appreciated that there are other applications where it is important to know the amount of rotation or translation of a camera is involved between different frames of an object or scene. For example, there are navigational applications in which a camera mounted in a robot or on a vehicle takes successive frames of a scene as the robot or vehicle moves past a scene to determine its position and knowledge of the rotation or translation of the camera is important to such determination.

4. Brief Description of Drawings

FIG. 1 will be helpful in a preliminary discussion of epipolar geometry.

FIG. 2 illustrates rotational and translational components of the displacement in an epipolar line resulting from some camera motion.

FIGs. 3A & 3B illustrate errors in epipolar lines for inaccurate translation and inaccurate rotation, respectively.

FIG. 4 is a flow diagram of the basic procedure used in the invention.

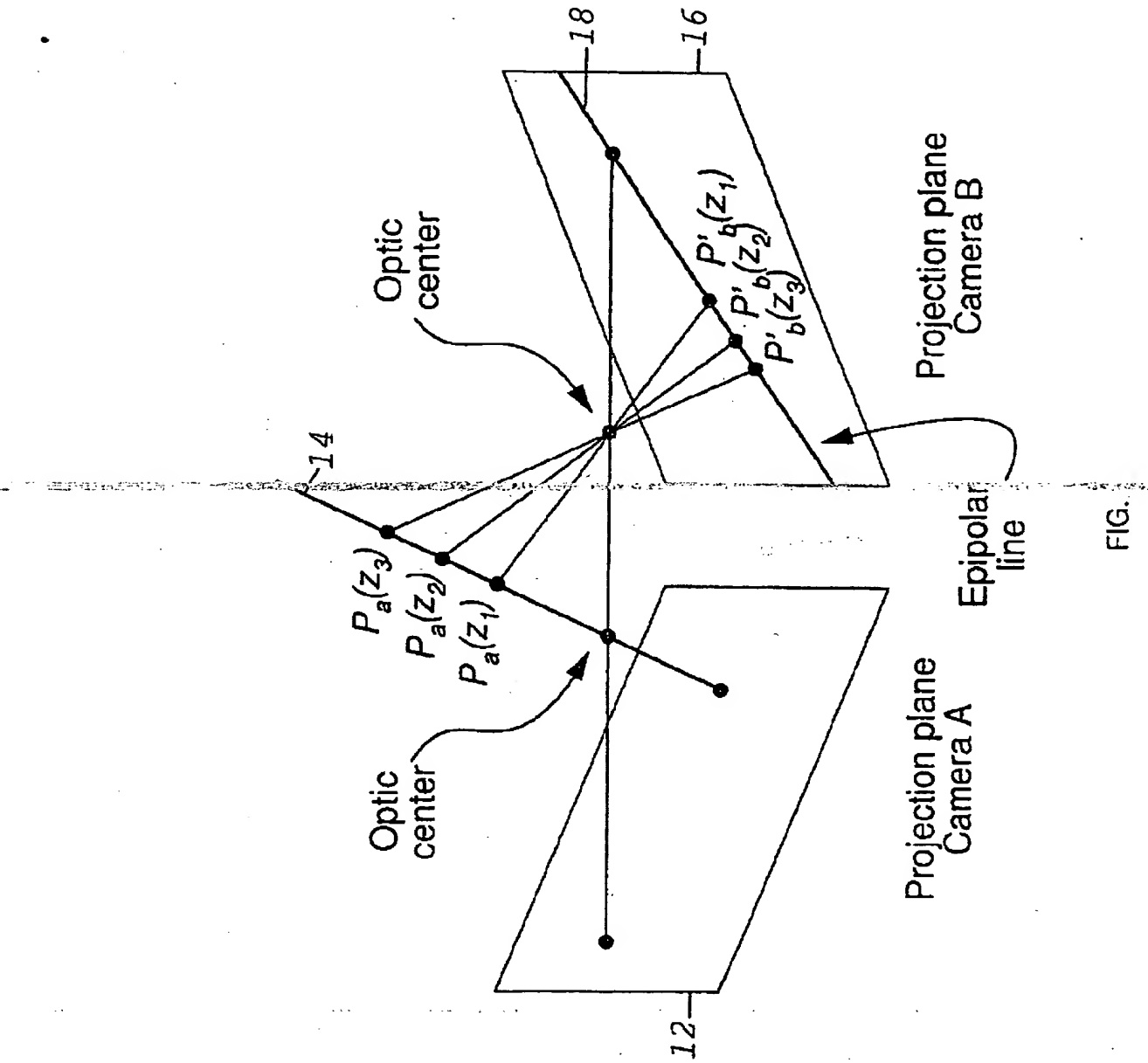

18
16
Optic center
$P'_b(z_1)$
$P'_b(z_2)$
$P'_b(z_3)$
Projection plane Camera B
14
$P_a(z_3)$
$P_a(z_2)$
$P_a(z_1)$
Epipolar line
Optic center
Projection plane Camera A
12

FIG.

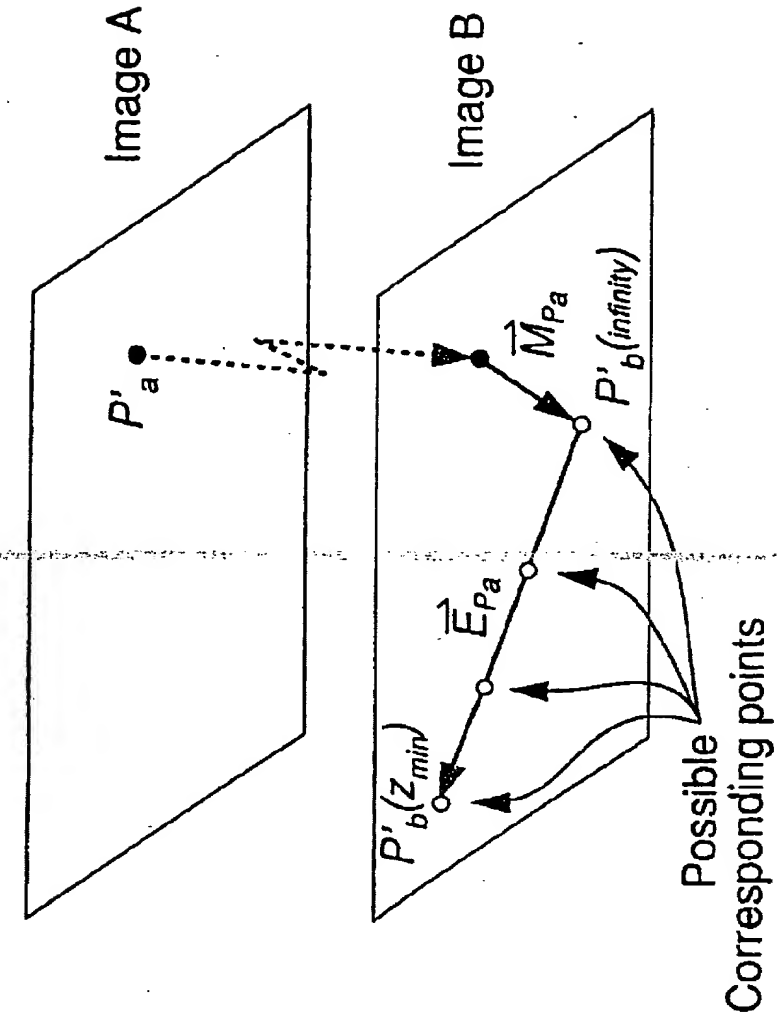


Image A
Image B
P'_a
$\vec{M}_{Pa}$
$P'_b(infinity)$
$\vec{E}_{Pa}$
$P'_b(z_{min})$
Possible
Corresponding points

FIG. 2

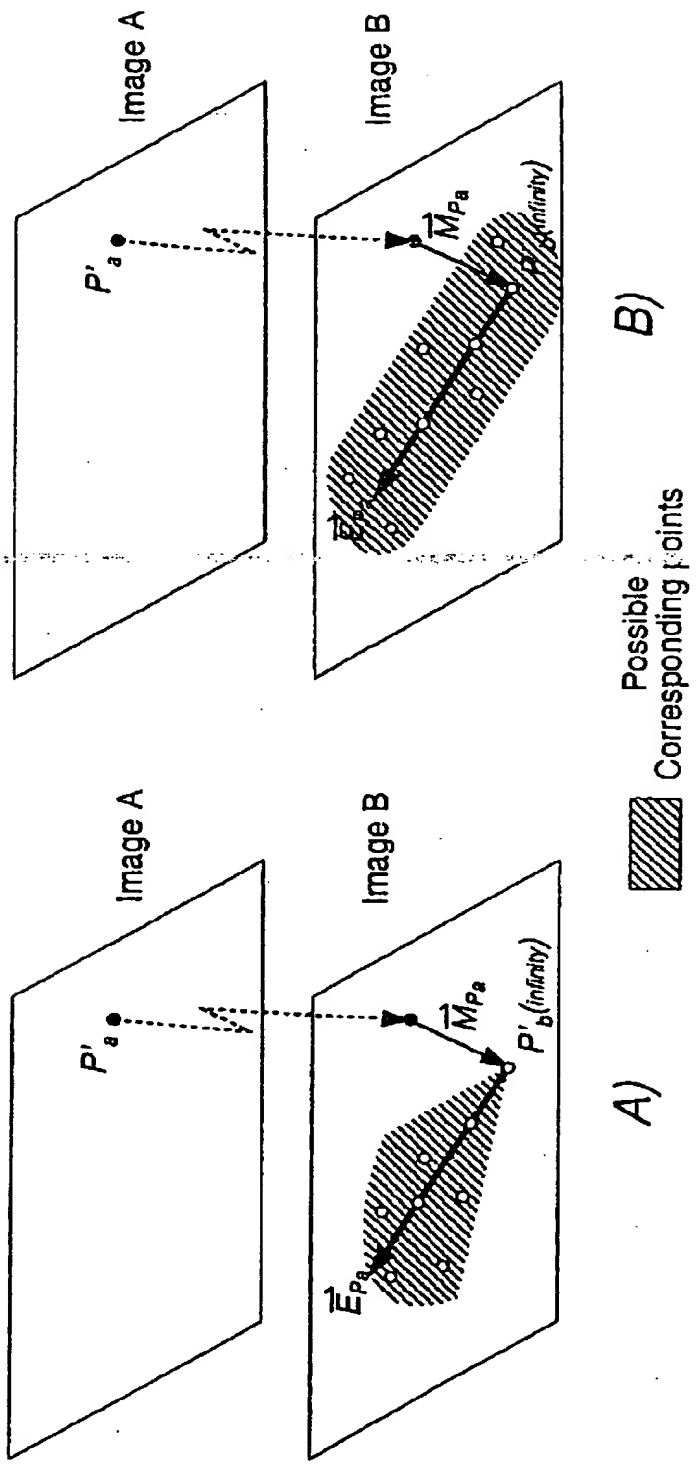


Image A
Image B
P'_a
$\vec{M}_{Pa}$
$P'_{b(infinity)}$
$\vec{E}_{Pa}$
B)
Possible
Corresponding points
A)
Image A
Image B
P'_a
$\vec{M}_{Pa}$
$P'_{b(infinity)}$
$\vec{E}_{Pa}$

FIG. 3

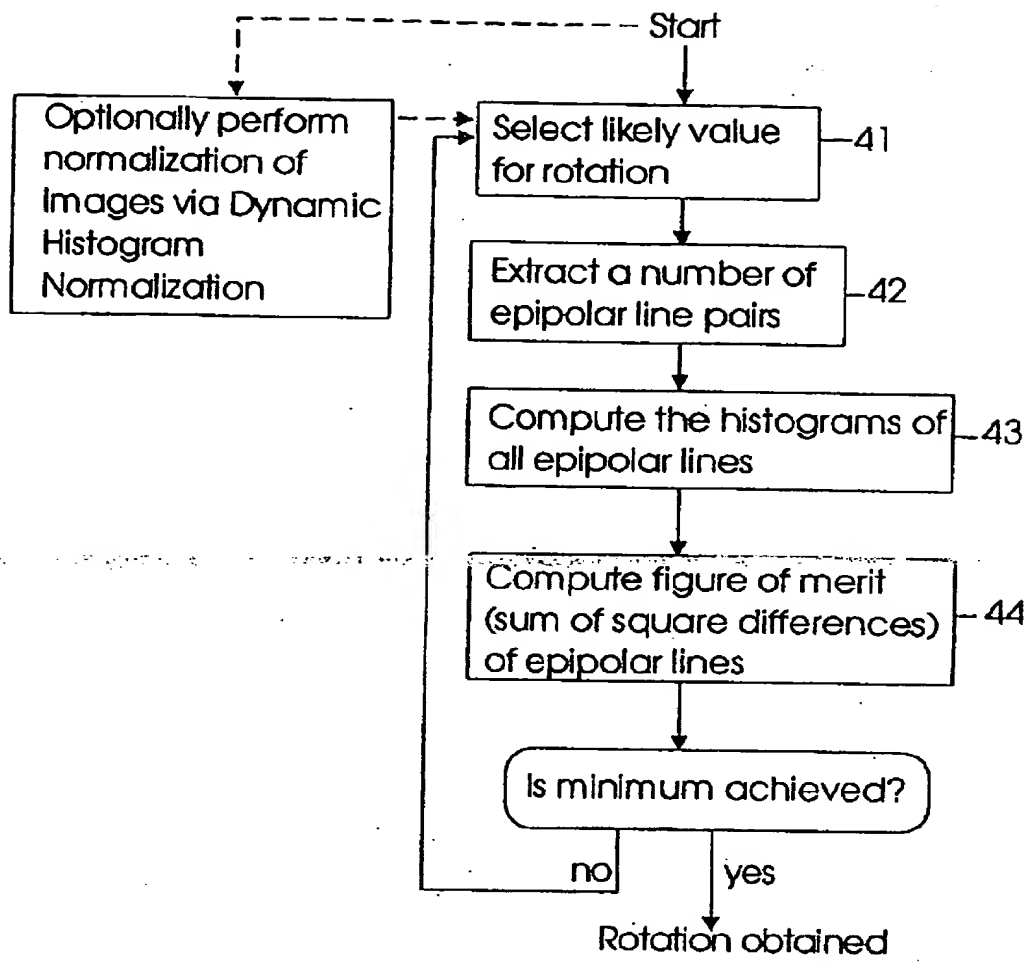


FIG. 4

1. Abstract

A technique for compensating for egomotion of the camera used to record a pair of two-dimensional views of a scene when the pair of images is to be used to provide a three dimensional representation of the scene. The technique involves comparing histograms of the intensity, levels of pixels of corresponding epipolar lines in the pair of images for assumed amounts of egomotion to identify the amoung that results in the smallest total of the sums of squared differences of the histograms.

2. Representative Drawing

FIG. 1